Evaluation of the impact of Adversarial Attacks using Face Filter Mobile Apps on Photoplethysmography (PPG) signals in estimating heart-rate in deep learning systems

Submitted as Research Report Master Dissertation in SIT723

27 February 2023

T3-2022

HyunDong Kim

STUDENT ID 221134523

COURSE - Master of Data Science (S777)

Supervised by: Dr. Lei Pan

Learning Outcomes: The focus of this paper is to examine the achievement of the Learning Outcomes in SIT723 unit (Appendix A) and Deakin Graduate Learning Outcomes (Appendix B) by reflecting on the research project titled "Evaluation of the impact of Adversarial Attacks using Face Filter Mobile Apps on Photoplethysmography (PPG) signals in estimating heart rate in deep learning systems".

Acknowledgements: Hyun Dong (Chris) Kim, the author of this paper, expresses gratitude to Dr. Lei Pan for his supervision and valuable feedback that helped to enhance the quality of this document. Dr. Lei Pan's guidance was instrumental in ensuring the comprehensiveness of the paper. It is worth noting that this research is a component of Deakin University Research Project A.

Abstract

This paper proposes an evaluation of the impact of the adversarial attacks on PPG-based systems for estimating heart rate in deep learning. Face Filter Mobile Apps (FFMA) and the Fast Gradient Sign Method (FGSM) are used to carry out the evaluation. The PPG signal data used in the evaluation were captured directly from the wrists of 7 participants. The data was collected in the presence of motion artifacts, and three different motions were included: resting, squatting, and stepping. To observe the changes in Human Activity Recognition (HAR) estimations, we conducted an experiment with FGSM adversarial attacks. The original dataset provided an optimal result for evaluating adversarial attack impacts. The experiment results showed a significant impact on HAR estimations when FGSM adversarial attacks were applied during the training phase. The accuracy dropped significantly for squatting and stepping, from 0.38 to 0.11 and 0.50 to 0.27, respectively. Accuracy for resting was unchanged. Further study is required to evaluate the overall impacts of using different adversarial attacks on PPG-based systems. This study should include FFMA on video-based datasets tested on different machine learning systems.
Keywords: photoplethysmography (PPG), adversarial attack, adversarial example, deep learning, heart rate estimation, human activity recognition, face filter apps, machine learning

Contents

List of Figures

List of Tables

1 Introduction

In recent years, there has been a growing interest in monitoring physiological signals that provide important information about an individual's health and wellbeing. A non-invasive and cost friendly technique, such as photoplethysmography (PPG), has been widely used for estimating physiological signals including heart rate [**1**, **2**, **3**]. PPG signals are obtained optically by shining light through the skin and detecting changes in light reflection in correspondence to the pulsation of underlying blood vessels have been widely studied and adapted in clinical practices.

For instance, in Human Activity Recognition (HAR), PPG signals are widely adapted as they are used to monitor physiological activity and estimating various physiological parameters such as heart rate, heart rate variability and oxygen saturation in real-time simply by using wearable devices on a person's wrist. The obtained PPG signals can then be analysed using machine learning algorithms like Recurrent Neural Network (RNN) [4, 5]to classify and identify patterns of different types of physical activities such as resting, squat, and stepping in the experiment. To evaluate the robustness of the model, a white-box attack, Fast Gradient Sign Method (FGSM), is used for comparing the accuracy of the model on clean and perturbed signals to showcase the vulnerability [**6**]. To statistically present the vulnerability, confusion matrix is adapted, and this clearly defines adversarial attacks do carry potentially high level of risks on monitoring heart rate for cardiac patients.

In addition, with the growing popularity of using face filter mobile apps has added a new dimension to the challenges faced by PPG-based heart rate estimation [1]. Face filters can mask and alter true characteristics of one's facial appearance such as skin colour, texture, shape and other characteristics, and thereby affect the quality and validity of PPG signals. The impact of face filters on PPG signals and heart rate estimation has not been fully explored, and there is a need to assess the vulnerability of PPG-based heart rate estimation to adversarial attacks using face filters.

Hence, the main objectives of this study are: (1) To investigate and evaluate on the vulnerability of adversarial attacks on PPG-based machine learning models (FGSM adversarial attacks on RNN model) (2) To design and implement adversarial attacks on PPG-based heart-rate estimation using face filters (using face filters as a source of adversarial attacks on a CNN model – future work) (3) To evaluate the performance of PPG-based heart rate estimation under both adversarial attack scenarios.

The findings of this study will contribute to the development of more robust and secured PPG-based heart rate estimation system.

2 Literature Review

In this section, an in-depth analysis of the state-of-the-art of PPG-based adversarial attacks in machine learning models as well as possible defences with a summary of examples of related works is discussed.

2.1 State-of-the-art research

2.1.1 Physiological signals

Physiological signals play a crucial role in detecting various illnesses including cardiovascular diseases [7]. With the advent of telehealth [8], which has expanded globally, especially after the COVID-19 pandemic, has made remote measurement of these signals increasingly important.

Table 1: Google Scholar academic publication number with keywords in title for Common Physiological Signals since 2016 as of 23 February 2023.

Common Signal	**Keywords**	**No. Publications since 2016 (as of 23 Feb 2023)**
Electro-encephalogram	EEG, Electroencephalogram, Electroencephalography	54,100
Electrocorticogram	ECG, EKG, Electrocardiogram	48,900
Electromyogram	EMG, Electromyogram	22,100
Electrocorticogram	ECoG, Electrocorticogram, Electrocorticography	15,400
Electrococulogram	EOG, Electroculogram	2,800
Electrodermal Activity	EDA, GSR, EDR, Electrodermal, Galvanic Skin Response	15,300
Photoplethysmogram	PPG, Photoplethysmogram, Photoplethysmography	7,650
Heart Rate Variability	HRV, Heart Rate Variability	18,300
Blood Pressure	Blood pressure	378,000
Oxygen Saturation	SpO2, Oxygen Saturation, Blood Oxygen	69,400
Pulse Rate	Pulse Rate	128,000
Respiration	Respiration	95,300
Skin Temperature	Skin Temperature	163,000
Eye Movement	Eye movement	101,000

In the studies carried by Wu et al. [7], illustrated the various physiological signals commonly used to compute measurements of human physiological processes, as shown in Figure 1. These physiological signals have been extensively studied in a range of applications, including clinical diagnosis and the development of wearable

devices for health monitoring and human-machine interactions, as indicated in Table 1.

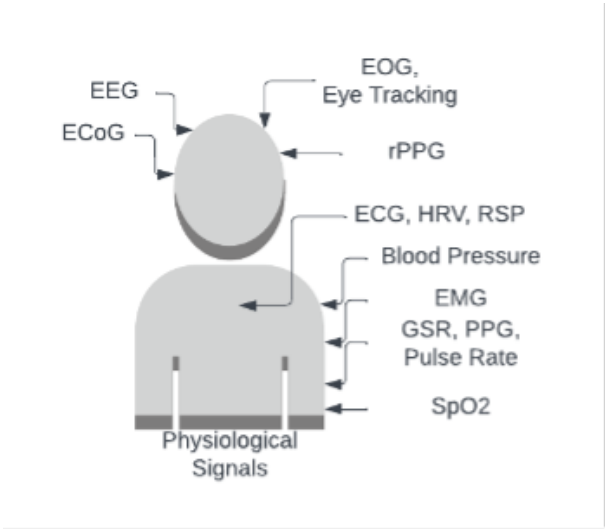

Figure 1: Common physiological signals and their measurement locations.

2.2 Photoplethysmography (PPG)

Photoplethysmography is a cost-effective optical method that uses a wearable heart-rate sensing device to detect changes in blood volume. The technique uses a light-emitting sensor placed on the skin, which emits light into the tissue [9]. A second sensor captures the amount of light that is reflected back to the device. This mechanism allows the device to collect information about the amount of emitted and reflected light from the patient's skin over time, which changes in response to changes in blood volume. This information is later used to estimate the patient's heart rate and blood pressure.

Even with the enhancement of wearable devices being much more efficient in size, better connectivity and mobility, multi-functional sensing platforms are still in the prototyping phase. Hence, researchers have been reporting results on cardiovascular multi-parameter monitoring with the solutions characterised by complex methods which consequently affects aspects such as mobility or autonomy [6, 10, 11]

To overcome this drawback, remote photoplethysmography (rPPG) was introduced. It only requires a camera to replace the wearable device on the patient's body, enabling the recording of physiological signals, including heart rate (HR), heart rate variability (HRV), respiratory rate (RR), and oxygen saturation (SpO2) [12].

2.3 Remote Photoplethysmography (rPPG)

Remote Photoplethysmography (rPPG) is a contactless measurement technique for physiological signals [9]. It does not require any physical contact with the patient and is accomplished by using a camera, which can be either a commercial or professional device, to capture several facial information, including changes in skin tone, lighting, and structure [10]. Similar to PPG, rPPG collects the difference between specular reflection (the pure light reflection from the skin) and diffused reflection (the remaining light reflection from absorption and scattering in skin tissue), which vary in response to changes in blood volume and are used to determine physiological signals.

Figure 2: rPPG Definition Diagram [10].

Proceudre of observing rPPG Signals is as below:

1. Position the face in front of the webcam to capture a video and use rPPG models to identify facial landmarks, which serve as a reference point for each participant. The region of interest is typically the two-thirds of the face where most blood vessels are concentrated.

2. Measure the average RGB colours of the participant's face over time, including both specular and diffused reflections.

3. Use the facial model to filter out any head motion noise and estimate a noise-free heart rate.

4. Analyse the pattern of peaks in the signals.

5. Estimate the participant's heart rate and heart rate variability.

2.4 Overview of adversarial attacks in machine learning

In machine learning, adversarial attacks refer to the methods used to manipulate or deceive the input of a machine learning model intentionally in a way that causes it to make incorrect decisions [11, 12]. These attacks compromise the reliability of the machine learning system by modifying input data, adding noise to input data, or even exploiting weaknesses in the model architecture or training process. The goal of adversarial attacks is often to cause the model to misclassify or mislabel the outcomes, which is severely harmful for medical machine learning practices [12].

2.5 Types of adversarial attacks

The adversarial attacks can broadly be categorised into two types:

2.5.1 White-box attack

In white-box attacks, the adversary has knowledge about the target model including its architecture, parameters, and gradients. Since the adversary knows everything about the model, it is the easiest attack scenario that maximises the harm to the model overall which often is the case that the model designer is carrying out an evaluation considering the model is under attack. Well-known white attacks include L-BFGS, Deep Fool, the CW method, the fast gradient sign method (FGSM) and many more [13].

2.5.2 Black-box attack

In black-box attacks, the adversary does not know about the architecture nor the parameters of the target model. Instead, the adversary has access to manipulate data inputs to the model to observe the outcomes. This is known to be the most realistic scenario in practice as the adversary can generate adversarial examples from one machine learning model and apply to fool another machine learning model. If it happens to solve somewhat similar task by both models, the adversary can easily manipulate the outcome of the targeted model by generating adversarial examples from the substitute model to attack the target model [14].

2.5.3 Different forms of adversarial attacks

In addition, adversarial attacks can take different forms, including but not limited to:

1. Image classification Perturbing an image to mislead a deep learning model into recognising an object as another.

2. Text classification Perturbing a text to mislead a natural language processing model into assigning it a different label.

3. Speech recognition Adding noise to an audio signal to deceive a speech recognition system into transcribing it incorrectly.

4. Anomaly detection Perturbing a time-series or a feature vector to hide an anomaly from a machine learning model designed to detect it.

For the scope of this article, anomaly detection using FGSM is discussed in detail with a clear demonstration of how the attacks have caused accuracy and reliability of the machine learning model.

2.6 Explanation of Fast Gradient Sign Method (FGSM) attack

The Fast Gradient Sign Method (FGSM) is a type of white-box attack that is popular and simple to use and is widely adapted for generating adversarial examples in machine learning [6]. It works by making small perturbations to the input features of a target model by changing the gradient of the loss function to maximise the loss to cause incorrect predictions.

This attack takes advantage of the gradients of the model's loss function with respect to the input features to maximise the loss by adding a small constant, called as the step size, to ensure that the adversarial example is not too far from the original input. This is then passed to the model, causing the model to make wrong predictions.

Figure 3 clearly represents how FGSM attack can have a significant impact on the test label classification results. This diagram illustrates that due to the gradient changes, Stepper could be classified as Squat and vice versa with such attacks. This article has adapted FGSM attacks due to its simplicity in analysing and being a baseline for evaluating the robustness of the target model against such attacks [16].

Figure 3: FGSM attack demonstration [15].

The below FGSM attack procedure is executed in the experiments and detailed analysis will be discussed.

1. Given an input example and a target model, adversary calculates the gradient of loss function with respect to input features.

2. The adversary then adds a small multiple of the gradient (Epsilon) to the input features, with the multiple chosen such that the magnitude of the adversarial perturbation is within a specified constraint.

3. The adversarial example is then passed to the target model, and the model's prediction is evaluated.

2.7 Defences

There are number of defences against adversarial attacks [6, 17] which can be applied in different stages.

Table 2: Defence against adversarial attacks

Defences	Examples	Refs
1. Data Modification	adversarial training, data compression, data randomisation, transferability blocking, and gradient hiding	[22] [23] [24] [25]
2. Model Modification	regularisation, deep contractive network, mask layer	[26-28]
3. Auxiliary Tools	adversarial detection models, defence-generative adversarial nets (defence-GAN)	[29, 30]

2.7.1 Data Modification

There are several ways of defending against adversarial attacks using data modification techniques. Adversarial training is one of the most popular adversarial defence approaches where the target model is trained with adversarial examples either during the training phase or applied in input data in testing phase. According to the experiment proposed by Hussein et al 2020 [27], the deep learning model was trained with adversarial samples for robust prediction of epilepsy seizures. Their experiments focused on overcoming the challenges of EEG-based seizure classification, including individual symptom differences as well as a shortage of pre-ictal labelled data. The idea behind it was to first construct a deep learning classifier model with EEG data and then perform white-box attacks on the classifier to obtain adversarial examples, which were then combined with the original data for retraining the model. The results have shown that adversarial training has increased both accuracy and classifier's robustness.

On the other hand, other researchers argue that adversarial training does improve deep learning neural network's robustness at the cost of undesirable accuracy degradation, as illustrated in Karim et al's work [28].

2.7.2 Model Modification

Sadehi et al. [29] proposed a regularisation-based model modification technique that aims to defend against adversarial attacks by optimising classifier parameters for accuracy and security using an analytical framework. The optimisation problem considers both test accuracy and robustness against adversarial attacks to determine the optimal classifier parameters. Experiments conducted by Sadehi et al. [29] on EEG-based eye state recognition (open or close) showed that this technique achieved high classification accuracy and robustness against black-box targeted evasion attacks, indicating its potential effectiveness as a defence mechanism. However, it is important to note that these model modification approaches often lack theoretical guarantees and could still be vulnerable to model-agnostic black-box attacks, which limits their reliability and highlights the need for further research to develop more robust defences against adversarial attacks.

2.7.3 Auxiliary Tools

Adversarial detection is a defence mechanism that employs a separate module to identify and take action against adversarial attacks. One of the simplest approaches involves discarding adversarial examples outright. In the case of ECG biometric authentication systems, Cai and Venkatasubramanian [30] proposed an approach to detect medical image adversarial attacks on DNNs from carefully-engineered signal perturbation of injection-based evasion attacks, which can cause inconsistencies in multiple physiological signals based on the same underlying physiological process. This approach demonstrated over 90 percent accuracy in detecting subtle ECG morphological alterations for both healthy subjects and patients. Similarly, Karimian et al. [31] proposed two strategies namely cross-subject attack and cross-device attacks to protect ECG biometric authentication systems from spoofing by evaluating whether the ECG signal characteristics match the corresponding heart rate variability or PPG features, similar to Cai and Venkatasubramanian's approach. However, adversarial detection relies heavily on distinguishing between adversarial and benign examples and may be ineffective against adaptive attacks [32] that can fool not only the classifier but also the detector.

3 Research Design & Methodology

3.1 Research Direction

This research investigation seeks to ascertain the extent to which a facial filter app impacts heart-rate estimation using PPG signals. Its objective is to propose a novel analysis of diverse adversarial attacks on both [33, 34] and RNN [5, 35] networks, assessing their effectiveness through various evaluation techniques.

This study relies on a pre-established and publicly available PPG dataset prior to exploring the development of facial filter adversarial attacks on a video-based dataset. By considering baseline adversarial attacks such as FGSM and data modification techniques [15], which have demonstrated accessibility and reliability

in previous research studies involve EEG and ECG, the designed model in this paper will be able to optimise the use of the dataset.

Moving forward, it is aimed to develop a more sophisticated CNN model on a video-based dataset such as MIMIC III by utilising both data and model modification techniques, which will be trained with facial filter apps to evaluate the robustness and reliability of the newly developed model. Ultimately, this study will also touch base on defence studies illustrated in Li et al.'s paper [36].

3.2 Research Approvals

Although the datasets contain personal information, an application for ethics review through the DUHREC (Deakin University Human Research Ethics Committee) is not required as they are open source. However, it is still important to handle the research with caution when accessing and analysing the data to protect the privacy of individuals. For adversarial studies on medical research, it is essential to disclose the use of privacy, and an isolated local environment is strongly recommended in line with Deakin University guidelines. This measure helps ensure that the data is secured, and privacy is not compromised. It is pleased to report that an approval from PhysioNet, MIT Laboratory for Computational Physiology, Institute for Medical Engineering and Science, was obtained to use the MIMIC III dataset for developing research models (see Appendix C). As a responsible researcher, will adhere to the terms of use and take all necessary precautions to maintain the confidentiality of personal information contained in the dataset.

3.3 Research Gap Analysis

The number of published studies on EEG and ECG exceeds those on PPG as shown in Table 1. Given the growing interest in PPG and its susceptibility to adversarial attacks, there is a tenacious need for research into adversarial machine learning. While some established papers offer valuable insights into PPG, there is a paucity of research on adversarial machine learning in this field. The use of face filter apps as a form of data modification is a feasible and easily executed adversarial technique for replicating white-box attacks and assessing model robustness and reliability [27]. Additionally, defence strategy studies have demonstrated that models can exhibit

higher levels of robustness when using data modification technique [11, 26, 37]. Consequently, employing the FGSM on a RNN in conjunction with face filter apps to introduce noise to the video-based CNN model for classification questions [6, 38]and comparing the results would provide a more comprehensive understanding of the vulnerability of a specific machine learning model as well as the quantitative impact of such an attack.

3.4 Research Questions

The research questions are listed as below:

• How to build a PPG signal based Recurrent Neural Network (RNN) to estimate Human Activity Recognition?

• How to deploy Fast Gradient Sign Method (FGSM), white-box adversarial attack on the target model?

• How does FGSM adversarial training affect the performance of the target model in determining HAR? (Quantitative Analysis)

4 Artefact Development Approach

The following list shows resources used to create such artefacts to address the research questions in T3 2022 and will be continued in T1 2023.

Table 3: Artefact Development

Artefacts Type	Description	Use	Ref
Hardware	MacBook Air M2, 8GB Memory	A hardware device to carry out experiments	-
Operating and System Software	MacOS Ventura 13.0.1 Google Colaboratory	An operating system in which experiment parameters to be tuned specifically to MacOS language. Commonly used machine learning system	-
Programming language	Python	Open-source programming language (Used in Google Colaboratory)	-
Networks	Recurrent Neural Network (RNN) Convolutional Neural Network (CNN) A mix of RNN and CNN	Open-source dynamic neural networks used for conducting experiments	[26, 29, 30, 39]
Algorithms	Tf.keras.models – based on open source code [26]	Creating models based on Tensorflow and Keras	
Requirements	TensorFlow Keras 2.9.0 TensorFlow 2.9.2	Pre-requisites to replicate the output of the experiment using open-source resources	
Datasets	MIMIC III PPG Signal data	Open-source dataset given that consent is approved (For future study) Raw PPG signal data used in the experiment; normalised for the purpose of the experiment	[19, 43]
Adversarial Attack	Fast Gradient Sign Method (FGSM)	A subject matter for testing machine learning system developed above	[6, 19]
Metric	Confusion Matrix	A mathematical model to statistically define the performance of a classification algorithm	[23]

As discussed in Wu et al., 2021 [7], it is proven by many researchers in the recent years that the PPG has gained popularity as the number of publications on PPG

has more than doubled in the past two years. However, it is still relatively less than topics discussed on EEG or ECG in adversarial attacks. Hence, this paper highlights providing detailed analysis on the number of research questions including but not limited to the below.

1. Establish RNN machine learning model based on PPG signals in determining HAR activities [44].
2. Train the model and evaluate the accuracy using confusion matrix.
3. Apply adversarial attacks on the same training data and evaluate on the affected results.
4. Compare the result (Clean vs Adversarial) to see how it impacts on the overall estimating process.
5. Observe and evaluate on changes of confusion matrices by conducting parameter tuning (what parameters epochs).

4.1 System Design

The experiment was conducted with the system below:

• All experiments were carried out on M2 Macbook Air with 8GB memory, utilising free GPU option enabled on Google Collaboratory.

• RNN was developed with Tensorflow 2.4.1 and Keras 2.4.0, Numpy 1.21.6.

• The network and the related algorithms were developed on Google Collaboratory.

4.2 Dataset

The dataset used in the experiment consists of PPG signals from seven subjects aged between 20 and 50 years old, including three males and four females, who performed different motions: resting, squatting and stepping. The dataset was divided into five sets with a total duration of 17,201 seconds and 210 recording sessions. For training, data from 5 subjects were used, while data from 2 subjects were reserved for testing.

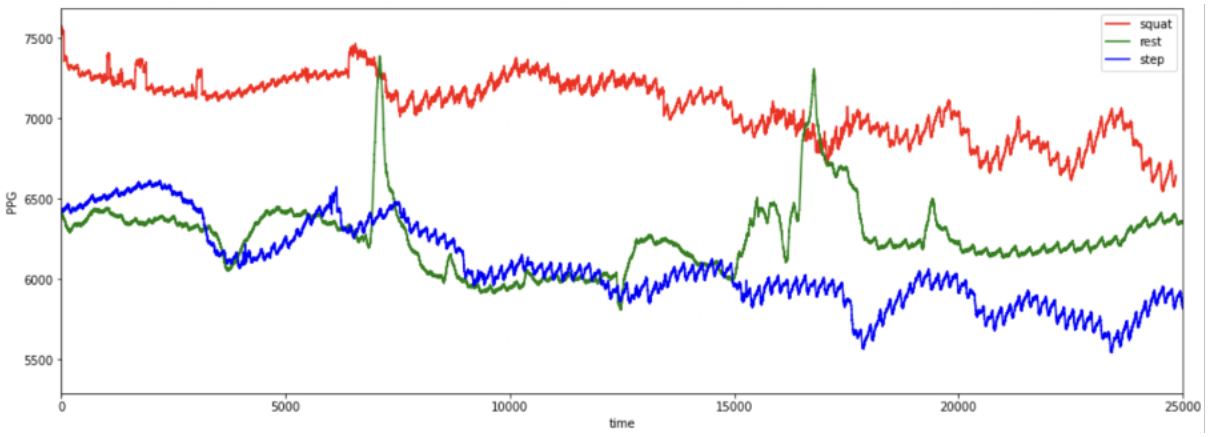

Figure 4: PPG Signals of subject ID1 on different physiological activities: Squat (Red), Rest (Green), and Stepper (Blue)

4.3 Experiment Design

The experiment consists of seven participants performing physiological activities for total of 17201 seconds with 210 samples as below.

i. Five series of ten squat exercises each.

ii. Five series of ten stepper exercises each.

iii. Five series of resting for five minutes each.

4.4 Data pre-processing and normalisation of PPG signals

Since the PPG signals of different physiological activities vary. Hence, the signals are normalised between 0 to 250,000 for better consistency [29, 45]

4.5 Design and implementation of RNN-based HAR models

RNN architecture

Recurrent Neural Network (RNN) is one of the most popular neural networks along with Convolutional Neural Network (CNN). It is designed to process sequential data where the order of data elements is important. This technique uses loops to maintain a state to process information from previous inputs, allowing it to determine patterns

and dependencies in sequential data such as text, speech, and time series data [4]. Besides, CNN is designed to process images and videos data using convolutional layers to extract features from the input data to reduce dimensionality of the data [37, 38]. It is commonly used in video and image classification, object detection, and image segmentation. Although CNN is a great fit for extracting RGB characteristics from a human face when extracting PPG signals, for this article, RNN is considered as the signals are rather extracted via wearable device than using a camera for time series analysis [5]. For future studies, both RNN and CNN will be considered.

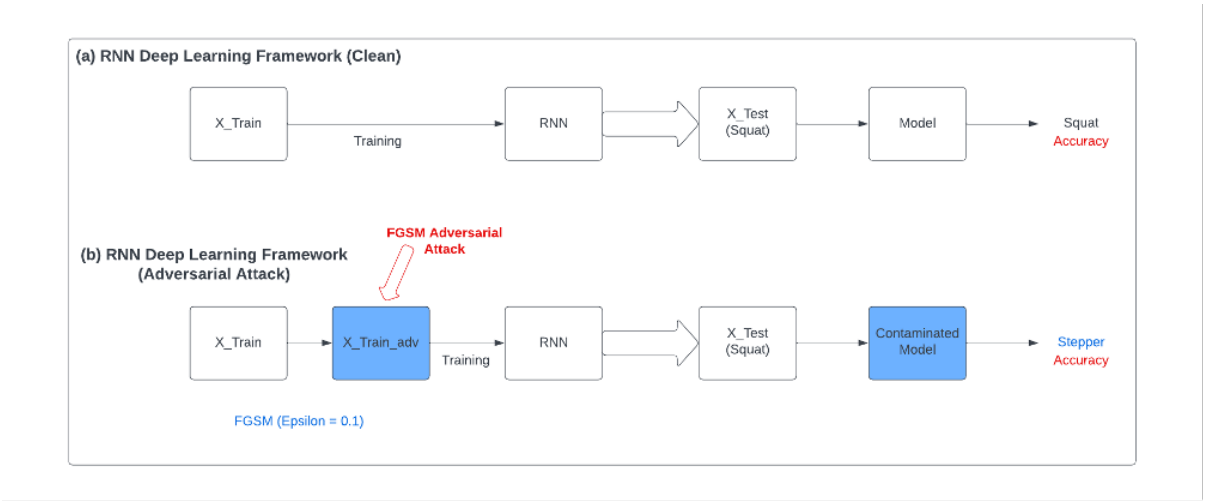

Figure 5: RNN Architecture

In the RNN architecture, target vectors from previous maps can be entirely based on historical information. This is different from traditional neural networks, which rely entirely on dependencies between adjacent layers.

The diagram in Figure 4 shows that the clean RNN model correctly identifies the label as ‘Squat’. However, when a FGSM adversarial attack is applied during the training stage, a contaminated model is created. In the contaminated model, the label is incorrectly classified as ‘Stepper’ instead of ‘Squat’. As a result, the adversarial attack significantly impairs the RNN model’s overall estimation performance.

The architecture of the RNN employed in this study is presented in Figure 5. It follows a commonly used framework for analysing time-based data (PPG data used in this experiment), comprising fully connected layers and LSTM cells.

```
### Permutation 1 Iteration 1
### creating model
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense1 (Dense)              (None, 600, 16)           80

 norm (BatchNormalization)   (None, 600, 16)           64

 lstm1 (LSTM)                (None, 600, 16)           2112

 drop2 (Dropout)             (None, 600, 16)           0

 lstm2 (LSTM)                (None, 600, 16)           2112

 drop3 (Dropout)             (None, 600, 16)           0

 lstm3 (LSTM)                (None, 16)                2112

 drop4 (Dropout)             (None, 16)                0

 dense2 (Dense)              (None, 3)                 51

=================================================================
Total params: 6,531
Trainable params: 6,499
Non-trainable params: 32
```

Figure 6: RNN Model in Google Colaboratory using a Sequential model algorithm

5 Research Evaluation

5.1 Project Deliverables

Project deliverables for Iteration 1 were achieved and partially for Iteration 2 during T3 2022. The rest of project deliverables are expected to be achieved in the following Trimester, T1 2023. A summary of project deliverables is stated below:

First iteration (T3 2022) - Completed

• Design and develop a Research Model based on Research Questions

• Set up working datasets (PPG Signals)

• Demo-run the model to see if all is up and running well

• Deploy white-box FGSM Adversarial Attacks

• Evaluate results using Confusion Matrix

• Produce a progressive report

Second iteration (T1 2023) – Partial Complete/Blocked (Technical issue found)

• Deploy white-box FGSM Adversarial Attacks

• Evaluate results using Confusion Matrix

• Produce a progressive report

• Extract PPG signal data from video-based MIMIC III dataset

• Define a new CNN Model based on research factors

• Complete technical verifications (Hardware issue found)

• Deploy white-box Face Filter App attack on the new model

• Analyse the results between RNN and CNN models

Third iteration (T1 2023) – Not Started

• Summarise Iteration 1 and Iteration 2 results

• Conclude answers to the research problems

• Produce a final thesis

5.2 Results of Research Deliverables (T3, 2022)

The results of FGSM adversarial attacks tested on RNN with GPU enabled on Google Colaboratory are discussed below with following parameters:

• Data augmentation applied (regularisation)

• Training and Testing dataset was split.

• Test performed on both clean and contaminated input data.

• A total of 100 training epochs

• Parameter variations on epsilon and test epochs

The accuracy comparison for FGSM adversarial attacks was evaluated using different levels of Epsilon and by tuning the model testing parameters. As shown in Figure 6, FGSM attacks on different parameters within the model significantly impact the estimation of HAR on a PPG signal based RNN model. When the RNN model was trained with adversarial examples, the HAR estimation for resting was unchanged, but the accuracy for both squat and stepper activities showed a significant difference. According to Table 3, the accuracy of the squat activity dropped significantly from 0.38 to 0.11 when Epsilon and Test epoch were set to 0.5 and 5, respectively. Similarly, the accuracy of the stepper activity dramatically decreased when Epsilon was set to 0.1 and Test epoch was set to 10, with the accuracy dropping from 0.5 to 0.27 (Table 3, Figure 6). Similar pattern is observed with the mean and standard deviation of the models as shown in Figure 7 and 8. The model subjected to adversarial attacks exhibits a more significant degree of deviation when compared to the model trained with clean data. Thus, it is evident that adversarial attacks do exert a discernible influence on the overall performance of the model, albeit the extent of this influence may appear trivial owing to several other factors, such as the limited size of the dataset.

These results show that FGSM white-box adversarial attacks significantly affect the training phase of the model, as the adversarial examples generated a slightly modified version of the PPG signals, leading to misclassification mainly on recognising the activities for squat and stepper. Thus, FGSM white-box attacks are effective in falsifying the machine learning system used in the experiment.

Since the current experiment focuses on the basic mechanism of white-box adversarial attacks, future studies will investigate the implications of different types and forms of adversarial attacks using face filters on a video-based dataset (MIMIC III) in a CNN model [33, 34].

Table 4: Confusion Matrix (Accuracy measurement) of FGSM Adversarial Attacks on Target Model.

FGSM Adversarial Attacks Confusion Matrix - Accuracy							
	Clean	**Adversarial Attacks**					
	(Eps 0.1, Epoch 10)	**(Eps 0.1, Epoch 5)**	**(Eps 0.1, Epoch 10)**	**(Eps 0.5, Epoch 5)**	**(Eps 0.5, Epoch 10)**	**(Eps 1.0, Epoch 5)**	**(Eps 1.0, Epoch 10)**
Resting	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Squat	0.38	0.42	0.28	0.11	0.17	0.12	0.24
Stepper	0.50	0.33	0.27	0.33	0.43	0.36	0.34

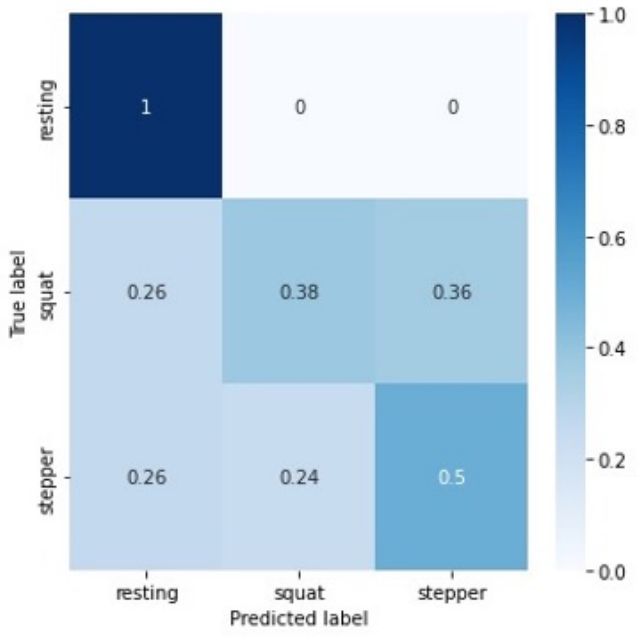

Figure 7: Clean (Train 100 epoch, Test 10 epoch, Epsilon 0.1)

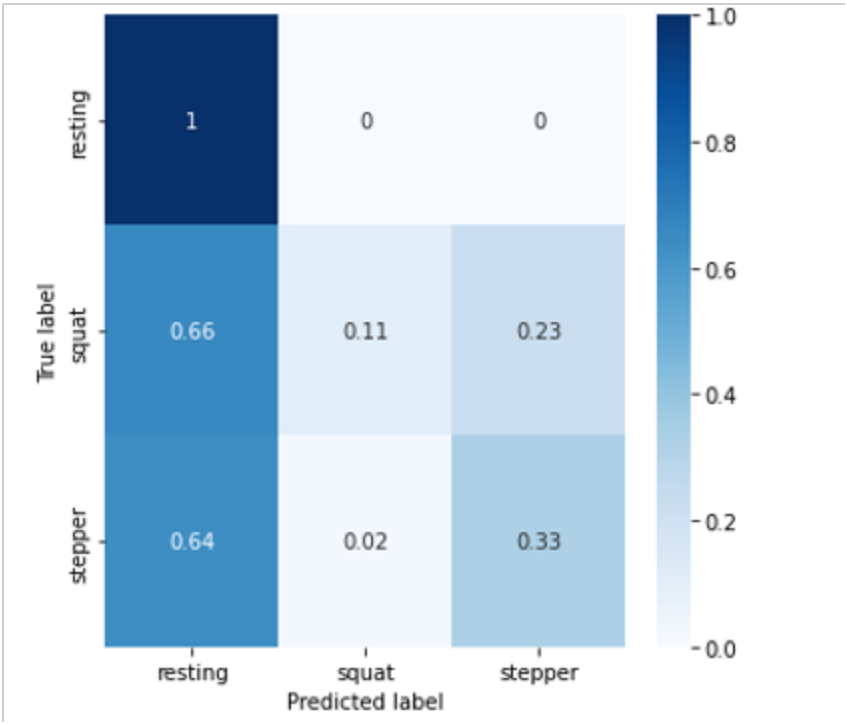

Figure 8: Adversarial attack (Test epoch 5, Epsilon 0.5)

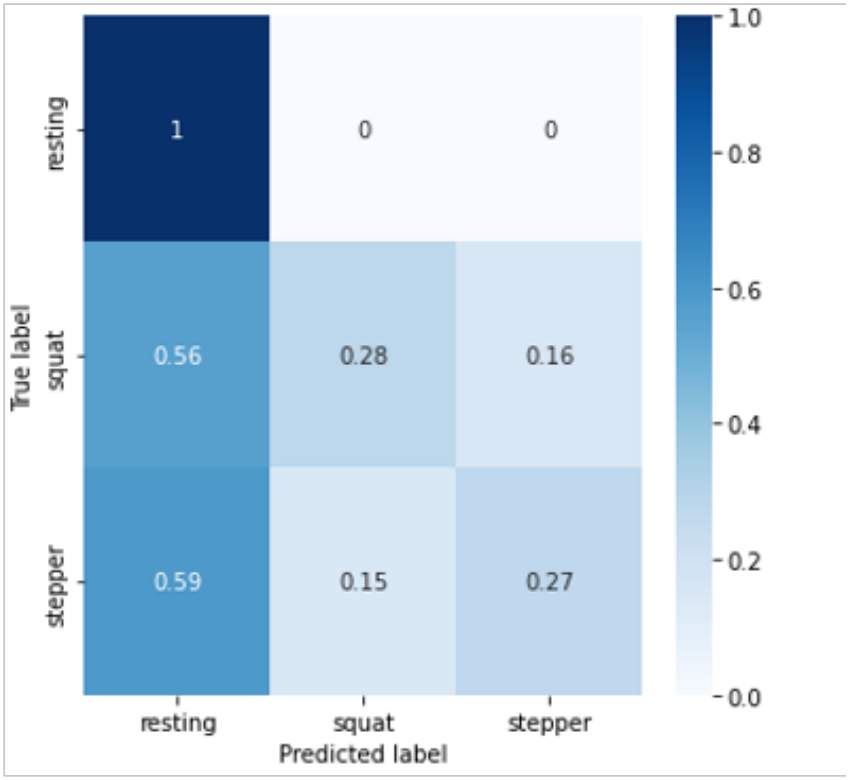

Figure 9: Adversarial attack (Test Epoch 10, Epsilon 0.1)

```
=================================================================
Total params: 1,731
Trainable params: 1,715
Non-trainable params: 16
_________________________________________________________________
Mean:  -2.8729382812233674e-16
Deviation:  1.0000000000000002
```

Figure 10: RNN Model in Google Colaboratory using a Sequential model algorithm

```
=================================================================
Total params: 1,731
Trainable params: 1,715
Non-trainable params: 16
_________________________________________________________________
Mean:  -0.0013451899
Deviation:  1.0363206
```

Figure 11: Mean and Standard Deviation of the model with clean data

5.3 Discussion on Research Deliverables (T3, 2022)

The experiment evaluated the impact of FGSM white-box adversarial attacks on a PPG signal-based RNN model used for HAR estimation. The results showed that the accuracy of the model was significantly affected by the adversarial examples, leading to misclassification mainly on recognising the activities for squat and stepper.

The effectiveness of the FGSM white-box attacks in falsifying the machine learning system highlights the importance of considering adversarial attacks as a potential threat to the models' robustness and reliability. Future studies should investigate the implications of different types and forms of adversarial attacks, including those using face filters on a video-based dataset, in a CNN model. Developing robust defence mechanisms against adversarial attacks is a critical area of research in machine learning and deep learning, and these findings underscore the need for continued research in this area. Overall, this study contributes to a better understanding of vulnerabilities and limitations machine learning models that highlights the importance of considering adversarial attacks in the development of evaluation of any machine learning models, particularly in healthcare applications where accurate measurements are critical for patient monitoring and diagnosis [42].

6 Project Management

6.1 Milestones

Table 4 represents the Gantt chart showing the key milestones achieved during T3 2022 and those planned for T1 2023. In the first half of T3 2022, our focus was on gaining a general understanding of the research topic and gathering in-depth knowledge. Thereafter, implemented sprints on a fortnightly basis until the end of T3 2022. To track the project's progress, the author developed an Excel spreadsheet accessible through the URL provided (https://www.overleaf.com/project/63fb6a4612cf284ea424ffa5).

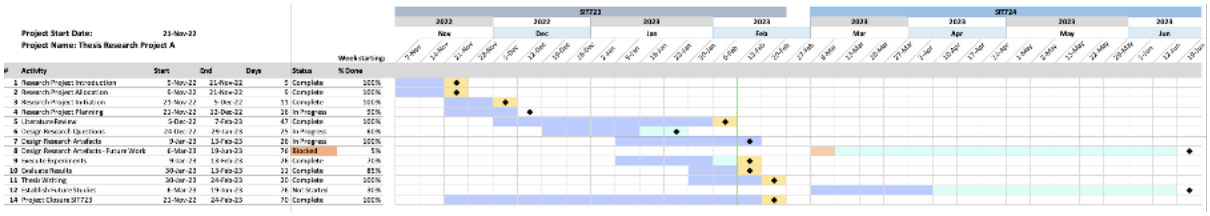

Figure 12: Project Research - Gantt Chart

Release 1 includes the following T3 2022 milestones from 9 November 2022 to 27

February 2023. In T3, the below milestones have been achieved. • Design of a RNN system

• Setup artefact developments

• Iteration 1: establishing a working model - completed.

• Iteration 2: adversarial attacks applied – completed.

• Results

• Evaluations

• Identify future study.

During this trimester, significant milestones have been achieved that provide evidence of the vulnerability of the PPG-based RNN model in Prototype Iteration 2, as presented in this work. The findings indicate that white-box attacks using FGSM significantly impacted the HAR estimation for both squat and stepper activities. This vulnerability was due to the characteristics of the observed PPG signals for these activities, which were much more fluctuating than those during rest, making them more susceptible to adversarial attacks, as shown in Figure 6. Despite the various types of adversarial attacks available, only FGSM was used in Prototype Model 1 during Iteration 2 to get the model up and running. Due to technical difficulties, some machine learning model parameters were not supported on Apple's silicon chipset, and therefore, applying facial filter mobile apps as part of a white-box attack on the MIMIC III dataset was postponed to the next trimester.

6.2 Project Logbook

As per Appendix D, a comprehensive project logbook demonstrates the exact course of action taken throughout the entire lifespan of the project. This not only highlights the number of days or weeks dedicated to completing significant project milestones, but also provides valuable insights for future endeavours. By analysing the contents of the logbook, time management can be optimised which will also lead to accurately forecasting the total duration of similar tasks in upcoming projects (https://www.overleaf.com/project/63fb6a4612cf284ea424ffa5).

6.3 Meeting notes with project details

Appendix E illustrates the recording of meeting logs and minutes which encapsulate crucial details in dot points pertaining to the project's advancement. While the entries in the logs maintain concision, they serve as a guide for ensuring the project's smooth progression as well as a checklist to facilitate continuous improvement (https://www.overleaf.com/project/63fb6a4612cf284ea424ffa5).

7 Risk Analysis Risk Management Plan

7.1 Risk Analysis

While this study provides valuable insights into the impact of FGSM white-box adversarial attacks on PPG signal-based RNN models for HAR estimation, there are several limitations to be considered.

Firstly, the sample size used in this study was relatively small, consisting of PPG signals from only seven subjects. This small sample size may limit the generalisability of the findings.

Secondly, this study focused on FGSM white-box attacks, which are only one type of adversarial attack. Future studies should investigate the impact of other types of adversarial attacks and their combinations on the video-based PPG signal dataset for estimating heart rate.

Lastly, the experiments were conducted on a relatively simple RNN model and did not explore more complex architectures, such as LSTM or GRU, which may be more robust to adversarial attacks.

7.2 Risk Management Plan

The risk management plan aims to mitigate the risks that have been identified in the Risks Analysis section. To this end, future experiments will deploy a more diverse and extensive dataset such as MIMIC III and leverage additional adversarial attacks including the application of face filter apps in combination with image classification techniques. Furthermore, a more sophisticated network architecture will be employed by utilising CNN to convert video-based signals into PPG signals, followed by running tests on RNN network to observe outcomes for further evaluations. The details of this concise risk management plan can be accessed through the URL provided (https://www.overleaf.com/project/6382384a076a3ff6d71c2dc5).

8 Conclusion & Future Work

8.1 Conclusion

This study provides complelling evidence that FGSM white-box adversarial attacks can significantly impact the accuracy of a RNN model in estimating human activity recognition based on PPG signals. By testing different parameter variations on epsilon and test epochs, the study found that adversarial attacks had a notable effect on the model's accuracy, especially for squat and stepper activities. Furthermore, the study revealed that the model trained with adversarial examples exhibited more significant deviation compared to the model trained with clean data. These findings have brought an attention to the implications for the development of robust and reliable machine learning models that can withstand adversarial attacks.

8.2 Future Work

In the light of the results, future studies should investigate the effects of different

types and forms of adversarial attacks on machine learning models, especially in a video-based dataset like MICIM III using more sophisticated models such as CNN in conjunction with RNN. This potentially leads to more exploration on the effects of facial filters which can significantly alter the appearance of participants in the dataset and lead to misclassification. By investigating the effects of adversarial attacks on different types of models, it is possible to gain a better understanding of the vulnerability of the medical machine learning system and develop more robust and reliable models that are better equipped to withstand adversarial attacks.

In addition, the study highlights the critical need to develop robust and reliable machine learning models and defensive techniques to ensure the continued progress and success of medical machine learning to take place in clinical applications.

9 Reference

[1] D. Biswas, N. Simões-Capela, C. V. Hoof, and N. V. Helleputte, "Heart Rate Estimation From Wrist-Worn Photoplethysmography: A Review," IEEE Sensors Journal, vol. 19, no. 16, pp. 6560-6570, 2019, doi: 10.1109/JSEN.2019.2914166.
[2] C. El-Hajj and P. A. Kyriacou, "A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure," Biomedical Signal Processing and Control, vol. 58, p. 101870, 2020/04/01/ 2020, doi: https://doi.org/10.1016/j.bspc.2020.101870.
[3] S. Yang, S. P. Morgan, S.-Y. Cho, R. Correia, L. Wen, and Y. Zhang, "Non-invasive cuff-less blood pressure machine learning algorithm using photoplethysmography and prior physiological data," (in eng), Blood Press Monit, vol. 26, no. 4, pp. 312-320, 2021/08// 2021, doi: 10.1097/mbp.0000000000000534.
[4] M. Alessandrini, G. Biagetti, P. Crippa, L. Falaschetti, and C. Turchetti, "Recurrent Neural Network for Human Activity Recognition in Embedded Systems Using PPG and Accelerometer Data," Electronics, vol. 10, no. 14, p. 1715, 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/14/1715.
[5] P. Su, X.-R. Ding, Y.-T. Zhang, J. Liu, F. Miao, and N. Zhao, "Long-term Blood Pressure Prediction with Deep Recurrent Neural Networks," p. arXiv:1705.04524doi: 10.48550/arXiv.1705.04524.
[6] R. Paul, M. Schabath, R. Gillies, L. Hall, and D. Goldgof, "Mitigating Adversarial Attacks on Medical Image Understanding Systems," in 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 3-7 April 2020 2020, pp. 1517-1521, doi: 10.1109/ISBI45749.2020.9098740.
[7] D. Wu et al., "Adversarial Attacks and Defenses in Physiological Computing: A Systematic Review," p. arXiv:2102.02729doi: 10.48550/arXiv.2102.02729.
[8] S. Omboni et al., "The worldwide impact of telemedicine during COVID-19: current evidence and recommendations for the future," Connected Health, vol. 1, 01/04 2022, doi: 10.20517/ch.2021.03.
[9] O. Nafea, W. Abdul, G. Muhammad, and M. Alsulaiman, "Sensor-based human activity recognition with spatio-temporal deep learning," Sensors, vol. 21, no. 6, p. 2141, 2021.
[10] L. Mirmohamadsadeghi, S. Fallet, V. Moser, F. Braun, and J.-M. Vesin, "Real-time respiratory rate estimation using imaging photoplethysmography inter-beat intervals," in 2016 Computing in Cardiology Conference (CinC), 2016: IEEE, pp. 861-

864.
[11] X. Yu, T. Laurentius, C. Bollheimer, S. Leonhardt, and C. H. Antink, "Noncontact monitoring of heart rate and heart rate variability in geriatric patients using photoplethysmography imaging," IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 5, pp. 1781-1792, 2020.
[12] G. Ribeiro, O. Postolache,
and F. F. Martín, "A Practical Approach to Health Status Monitoring Based on Heart Rate and Respiratory Rate Assessment," in 2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA), 22-24 June 2022 2022, pp. 1-6, doi: 10.1109/MeMeA54994.2022.9856576.
[13] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, and Z. J. Wang, "Video-Based Heart Rate Measurement: Recent Advances and Future Prospects," IEEE Transactions on Instrumentation and Measurement, vol. 68, no. 10, pp. 3600-3615, 2019, doi: 10.1109/TIM.2018.2879706.
[14] R. M. Seepers, W. Wang, G. d. Haan, I. Sourdis, and C. Strydis, "Attacks on Heartbeat-Based Security Using Remote Photoplethysmography," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 3, pp. 714-721, 2018, doi: 10.1109/JBHI.2017.2691282.
[15] N. Akhtar and A. Mian, "Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey," p. arXiv:1801.00553doi: 10.48550/arXiv.1801.00553.
[16] S. G. Finlayson, H. W. Chung, I. S. Kohane, and A. L. Beam, "Adversarial Attacks Against Medical Deep Learning Systems," p. arXiv:1804.05296doi: 10.48550/arXiv.1804.05296.
[17] X. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial Examples: Attacks and Defenses for Deep Learning," p. arXiv:1712.07107doi: 10.48550/arXiv.1712.07107.
[18] X. Sun and S. Sun, "Adversarial Attacks for Multi-view Deep Models," p. arXiv:2006.11004doi: 10.48550/arXiv.2006.11004.
[19] J. Kang. "Pytorch Adversarial Attack - FGSM/PGD." https://rainbow.tistory.com/entry/

[20] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and Harnessing Adversarial Examples," p. arXiv:1412.6572doi: 10.48550/arXiv.1412.6572.
[21] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. Berkay Celik, and A. Swami, "The Limitations of Deep Learning in Adversarial Settings," p. arXiv:1511.07528doi: 10.48550/arXiv.1511.07528.
[22] F. Tramèr, A. Kurakin, N. Papernot, I. Goodfellow, D. Boneh, and P. McDaniel, "Ensemble Adversarial Training: Attacks and Defenses," p. arXiv:1705.07204doi:

10.48550/arXiv.1705.07204.
[23] N. Das et al., "Keeping the Bad Guys Out: Protecting
and Vaccinating Deep Learning with JPEG Compression," p. arXiv:1705.02900doi: 10.48550/arXiv.1705.02900.
[24] C. Xie, J. Wang, Z. Zhang, Y. Zhou, L. Xie, and A. Yuille, "Adversarial Examples for Semantic Segmentation and Object Detection," p. arXiv:1703.08603doi: 10.48550/arXiv.1703.08603.
[25] H. Hosseini, Y. Chen, S. Kannan, B. Zhang, and R. Poovendran, "Blocking Transferability of Adversarial Examples in Black-Box Learning Systems," p. arXiv:1703.04318doi: 10.48550/arXiv.1703.04318.
[26] J. Gao, B. Wang, Z. Lin, W. Xu, and Y. Qi, "DeepCloak: Masking Deep Neural Network Models for Robustness Against Adversarial Samples," p. arXiv:1702.06763doi: 10.48550/arXiv.1702.06763.
[27] S. Maqsood, S. Xu, M. Springer, and R. Mohawesh, "A Benchmark Study of Machine Learning for Analysis of Signal Feature Extraction Techniques for Blood Pressure Estimation Using Photoplethysmography (PPG)," IEEE Access, vol. 9, pp. 138817-138833, 2021, doi: 10.1109/ACCESS.2021.3117969.
[28] H. Xiao, B. Biggio, B. Nelson, H. Xiao, C. Eckert, and F. Roli, "Support Vector Machines under Adversarial Label Contamination," p. arXiv:2206.00352doi: 10.48550/arXiv.2206.00352.
[29] K. Fariha Hossain et al., "ECG-Adv-GAN: Detecting ECG Adversarial Examples with Conditional Generative Adversarial Networks," p. arXiv:2107.07677doi: 10.48550/arXiv.2107.07677.
[30] P. Samangouei, M. Kabkab, and R. Chellappa, "Defense-GAN: Protecting Classifiers Against Adversarial Attacks Using Generative Models," p. arXiv:1805.06605doi: 10.48550/arXiv.1805.06605.
[31] A. Hussein, M. Djandji, R. A. Mahmoud, M. Dhaybi, and H. Hajj, "Augmenting DL with adversarial training for robust prediction of epilepsy seizures," ACM Transactions on Computing for Healthcare, vol. 1, no. 3, pp. 1-18, 2020.
[32] F. Karim, S. Majumdar, and H. Darabi, "Adversarial Attacks on Time Series," p. arXiv:1902.10755doi: 10.48550/arXiv.1902.10755.
[33] K. Sadeghi, A. Banerjee, and S. K. Gupta, "An analytical framework for security-tuning of artificial intelligence applications under attack," in 2019 IEEE International Conference On Artificial Intelligence Testing (AITest), 2019: IEEE, pp. 111-118.
[34] H. Cai and K. K. Venkatasubramanian, "Detecting Signal Injection Attack-Based Morphological Alterations of ECG Measurements," in 2016 International Conference on Distributed Computing in Sensor Systems (DCOSS), 26-28 May 2016 2016, pp.

127-135, doi: 10.1109/DCOSS.2016.36.
[35] N. Karimian, D. Woodard, and D. Forte, "ECG Biometric: Spoofing and Countermeasures," IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 2, no. 3, pp. 257-270, 2020, doi: 10.1109/TBIOM.2020.2992274.
[36] N. Carlini and D. Wagner, "Adversarial Examples Are Not Easily Detected: Bypassing Ten Detection Methods," p. arXiv:1705.07263doi: 10.48550/arXiv.1705.07263.
[37] R. Mutegeki and D. S. Han, "A CNN-LSTM approach to human activity recognition," in 2020 international conference on artificial intelligence in information and communication (ICAIIC), 2020: IEEE, pp. 362-366.
[38] K. Xia, J. Huang, and H. Wang, "LSTM-CNN architecture for human activity recognition," IEEE Access, vol. 8, pp. 56855-56866, 2020.
[39] S. W. Pienaar and R. Malekian, "Human Activity Recognition using LSTM-RNN Deep Neural Network Architecture," in 2019 IEEE 2nd Wireless Africa Conference (WAC), 18-20 Aug. 2019 2019, pp. 1-5, doi: 10.1109/AFRICA.2019.8843403.
[40] L. Li, C. Chen, L. Pan, J. Zhang, and Y. Xiang, "Video is All You Need: Attacking PPG-based Biometric Authentication," p. arXiv:2203.00928doi: 10.48550/arXiv.2203.00928.
[41] H. Zhang, Y. Yu, J. Jiao, E. P. Xing, L. El Ghaoui, and M. I. Jordan, "Theoretically Principled Trade-off between Robustness and Accuracy," p. arXiv:1901.08573doi: 10.48550/arXiv.1901.08573.
[42] K. Eykholt et al., "Robust Physical-World Attacks on Deep Learning Visual Classification," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18-23 June 2018 2018, pp. 1625-1634, doi: 10.1109/CVPR.2018.00175.
[43] T. Zebin, M. Sperrin, N. Peek, and A. J. Casson, "Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks," in 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC), 2018: IEEE, pp. 1-4.
[44] P. E. Novac et al., "Toward unsupervised Human Activity Recognition on Microcontroller Units," in 2020 23rd Euromicro Conference on Digital System Design (DSD), 26-28 Aug. 2020 2020, pp. 542-550, doi: 10.1109/DSD51259.2020.00090.
[45] G. Biagetti, P. Crippa, L. Falaschetti, L. Saraceni, A. Tiranti, and C. Turchetti, "Dataset from PPG wireless sensor for activity monitoring," Data in Brief, vol. 29, p. 105044, 2020/04/01/ 2020, doi: https://doi.org/10.1016/j.dib.2019.105044.
[46] A. Iqtidar Newaz, N. Imtiazul Haque, A. K. Sikder, M. Ashiqur Rahman, and A. Selcuk Uluagac, "Adversarial Attacks to Machine Learning-Based Smart Healthcare

Systems," p. arXiv:2010.03671doi: 10.48550/arXiv.2010.03671.

10 Appendix

10.1 Appendix A

ULO	These are the Learning Outcomes (ULO) for this unit. At the completion of this unit, successful students can:	Deakin Graduate Learning Outcomes
ULO1	Demonstrate the ability to explore an IT area in depth, define a research direction and scope, and develop a research proposal.	GLO1: Discipline-specific knowledge and capabilities GLO4: Critical thinking
ULO2	Demonstrate clear understanding of the broader context associated with a research project, including any safety, sustainability, and ethical considerations relevant to the project.	GLO8: Global citizenship
ULO3	Effectively and convincingly communicate and defend their understanding of technical content, objectives, design reasoning, and implementation details of a research projects in written and oral forms.	GLO1: Discipline-specific knowledge and capabilities GLO2: Communication GLO5: Problem solving GLO7: Teamwork
ULO4	Produce and discuss results to validate and/or justify the proposed project objectives and methodology.	GLO1: Discipline-specific knowledge and capabilities GLO3: Digital literacy
ULO5	Implement and deliver on a project's objectives by applying appropriate techniques, management practices, and associated skills and knowledge.	GLO1: Discipline-specific knowledge and capabilities GLO6: Self-management

Figure 13: Unit Learning Outcomes (ULO)

10.2 Appendix B

This task assesses your achievement of these Graduate Learning Outcome(s)	GLO1 – through the application of domain knowledge and expertise within the project. GLO2 – through engaging in presentations and thesis writing to communicate project outcomes and progress to a range of stakeholders. GLO3 – through the use of technology to locate and disseminate information related to their project GLO4 – through critical review of literature, work related to the project, and findings GLO5 – through working through project scoping and execution GLO6 – by working effectively to meet project goals and produce the required outputs GLO7 – by working with supervisors, other students, and members of associated research teams to help realise the project goals GLO8 – through documenting and discussing the broader context of the project and any associated complexities. In addition to the GLOs associated with the project work, students will also address all GLOs by collating and presenting evidence of the attainment of the course learning outcomes. This will incorporate work undertaken throughout the course.

Figure 14: Graduate Learning Outcomes (GLOs)

10.3 Appendix C

Dear Hyun Dong Kim,

Thank you for your interest in the PhysioNet Clinical Databases. We are pleased to say that your application for credentialed access has been approved.

You are now able to access protected databases upon agreeing to the terms of usage and completing any required training. For example, you can access MIMIC-III by following the steps below:

- Go to the project page at https://physionet.org/content/mimiciii/
- Find the "Files" section in the project description
- Follow instructions to complete required training, if necessary.
- Click "Sign the Data Use Agreement" to agree to the terms of usage for the dataset

To check the status of your training or to submit training documents for review, please visit https://physionet.org/settings/training/.

Regards, The PhysioNet Team, MIT Laboratory for Computational Physiology Institute for Medical Engineering and Science, MIT, E25-505 77 Massachusetts Ave, Cambridge, MA 02139

Figure 15: PhysioNet Application Approval

10.4 Appendix D

Student Name	Supervisor	Project Title	Date of Activity	Duration (Minutes)	Activity Type	Short Description
Name as registered	Primary Supervisor	Project Title	Must be in format 15/07/2021	Select from list	Select from list	Please fill out with succinct details
Hyun Dong Kim	Professor Lei Pan		15/11/2022	30	Supervision Meeting	
Hyun Dong Kim	Professor Lei Pan		18/11/2022	30	Supervision Meeting	
Hyun Dong Kim	Professor Lei Pan		25/11/2022	30	Supervision Meeting	
Hyun Dong Kim	Professor Lei Pan		16/11/2022	60	Literature Review	What is PPG
Hyun Dong Kim	Professor Lei Pan		17/11/2022	120	Literature Review	Recommended Paper Research
Hyun Dong Kim	Professor Lei Pan		17/11/2022	120	Literature Review	Evaluate the paper and start planning on my next goal
Hyun Dong Kim	Professor Lei Pan		27/11/2022	60	OnTrack Task	1.1P Ontrack Task
Hyun Dong Kim	Professor Lei Pan		27/11/2022	60	OnTrack Task	1.2P Ontrack Task
Hyun Dong Kim	Professor Lei Pan		27/11/2022	60	OnTrack Task	2.1P Ontrack Task
Hyun Dong Kim	Professor Lei Pan		27/11/2022	60	OnTrack Task	3.1P Ontrack Task
Hyun Dong Kim	Professor Lei Pan		3/12/2022	120	Up-skilling	Dataset research - MICIM (trying to get the quiz done but there's cost related to it)
Hyun Dong Kim	Professor Lei Pan		4/12/2022	90	Up-skilling	Dataset research - Kaggle (was looking for something similar like MICIM dataset)
Hyun Dong Kim	Professor Lei Pan		4/12/2022	60	OnTrack Task	4.1P Ontrack Task
Hyun Dong Kim	Professor Lei Pan		3/12/2022	90	Other	Week 4 Workshop / Lecture
Hyun Dong Kim	Professor Lei Pan		26/11/2022	120	Other	Week 2 Workshop / Lecture
Hyun Dong Kim	Professor Lei Pan		27/11/2022	120	Other	Week 3 Workshop / Lecture
Hyun Dong Kim	Professor Lei Pan		6/12/2022	120	Other	Week 5 Workshop / Lecture
Hyun Dong Kim	Professor Lei Pan		8/12/2022	120	Other	Week 6 Workshop / Lecture
Hyun Dong Kim	Professor Lei Pan		15/12/2022	120	Other	Week 7 Workshop / Lecture
Hyun Dong Kim	Professor Lei Pan		22/12/2022	120	Other	Week 8 Workshop / Lecture
Hyun Dong Kim	Professor Lei Pan		29/12/2022	120	Other	Week 9 Workshop / Lecture
Hyun Dong Kim	Professor Lei Pan		5/1/2023	240	Other	New research plan - research
Hyun Dong Kim	Professor Lei Pan		12/1/2023	240	Other	Trying new open source code found online
Hyun Dong Kim	Professor Lei Pan		19/1/2023	240	Other	Trying new open source code found online
Hyun Dong Kim	Professor Lei Pan		26/1/2023	240	Other	Established a new model
Hyun Dong Kim	Professor Lei Pan		2/2/2023	360	Other	Implemented adversarial attacks and collected results
Hyun Dong Kim	Professor Lei Pan		9/2/2023	660	Other	Thesis writing start
Hyun Dong Kim	Professor Lei Pan		16/2/2023	480	Other	Thesis writing continue
Hyun Dong Kim	Professor Lei Pan		23/2/2023	600	Other	Thesis writing review
Hyun Dong Kim	Professor Lei Pan		26/2/2023	600	Other	Thesis finalisation
Hyun Dong Kim	Professor Lei Pan		26/2/2023	200	Other	Presentation
Hyun Dong Kim	Professor Lei Pan		26/2/2023	200	OnTrack Task	Ontract Tasks

Figure 16: Project Logbook

10.5 Appendix E

Date	**24/2/2023**	
Attendees:	Chris	Dr. Lei Pan
Discussion		
1. Briefly discussed progress of thesis		
2. Received a brief guidance on what to touch on		
3. Dr Lei Pan suggested me to complete them asap		
4. Asked questions regarding presentation and ontrack tasks		
5. Also told to finish them ASAP		
6. Overleaf links to be updated		
Evidence		

Figure 17: Meeting minutes and logs